



Comparing upscaling algorithms from HD to ultra HD by evaluating preference of experience

Jing Li, Yao Koudota, Marcus Barkowsky, Hélène Primon, Patrick Le Callet

► To cite this version:

Jing Li, Yao Koudota, Marcus Barkowsky, Hélène Primon, Patrick Le Callet. Comparing upscaling algorithms from HD to ultra HD by evaluating preference of experience. The International Workshop on Quality of Multimedia Experience (QoMEX) 2014, Sep 2014, Singapore, Singapore. 10.1109/QoMEX.2014.6982320 . hal-01071465

HAL Id: hal-01071465

<https://hal.science/hal-01071465>

Submitted on 14 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMPARING UPSCALING ALGORITHMS FROM HD TO ULTRA HD BY EVALUATING PREFERENCE OF EXPERIENCE

Jing Li[†], Yao Koudota[†], Marcus Barkowsky[†], H       Primon[ ], Patrick Le Callet[ ]

[ ] LUNAM Universit  , Universit   de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes
rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France.

Email: {jing.li2, yao.koudota, marcus.barkowsky, patrick.lecallet}@univ-nantes.fr

[ ] VITEC, 99 rue Pierre S       92320 Chatillon, France

Email: helene.primon@vitec.com

ABSTRACT

As the next generation of TV, Ultra High Definition Television (UHDTV) is attracting more and more people’s attention as it provides a new viewing experience. Considering content delivery, due to the lack of Ultra HD resources, a direct question for the industry is that whether the state-of-the-art upscaling algorithms can be utilized to upscale the current HD or Full HD resources to UHD, gaining benefit from the higher resolution but without losing the high quality viewing experience. To investigate this, in this study, we upscaled 720p and 1080p sequences to UHD resolution by different upscaling algorithms. Paired Comparison methodology was used in the subjective experiment to evaluate their performances. The results showed that for the case of fast motion content, viewers’ preference on different upscaled video sequences is not significantly different. In general conditions, the low complexity upscaling algorithms (e.g., lanczos-3) performed better than the high complexity algorithms (e.g., Robust Super Resolution algorithm). A novel upscaling algorithm adapted to UHD is recommended to be developed based on the mechanisms of human visual system.

Index Terms— Ultra HD, Quality of Experience, Preference, Upscaling, Pair Comparison

1. INTRODUCTION

The UHDTV standard has emerged as the natural evolution of the current HD and 3DTV formats, which provide a realistic experience for the end-user in an immersive environment. The UHDTV standard defines two formats named 4K and 8K, with four and sixteen times of the HD resolution respectively. However, UHDTV is much more than higher spatial resolution, but involves a higher temporal resolution, an enhanced colorimetry, a higher bit depth, and a higher dynamic range than the previous HD format.

The human visual system reacts to the higher resolution and the larger screen size in conscious and subconscious ways. The observers’ viewing behavior and viewing experience may change compared to the HD or Full HD TV, notably concerning the viewing distance and the expectation on the perceived video quality. The perceptual quality is not only a matter of image quality, but also of immersiveness, naturalness and visual comfort. In accordance

with other advanced multimedia technologies, e.g., 3DTV, the term Quality of Experience (QoE) [1] is used to define this multi-dimensional “viewing experience”.

With the aim to achieve an enhanced end-user QoE, the broadcast industry is involved in the development of the end-to-end production and distribution chain based on the new UHDTV technology. For example, for content delivery, a first step might be the upscaling of the current HD video content to UHD.

Generally, upscaling can be implemented by image interpolation, which is achieved by curve fitting or regression analysis. “Bilinear”, “bicubic” [2] and “lanczos” [3] are the most available interpolation methods in most programs or softwares for resizing an image. Generally, “bilinear” has a smoothing effect, which will produce blurred but jagged edges. “Bicubic” would keep edges smooth and to some extent increase perceived contrast. Compared with the two above, “lanczos” is considered as the best algorithms in performance [3] [4] and it is the best compromise between reduction of aliasing, sharpness and minimal ringing.

Another way for upscaling is super resolution which is a resolution enhancement technic. Traditionally, it refers to the process that recovering a high resolution image by utilizing one or several different subpixel shifted low resolution images. However, the performance will highly depend on, for example, the number of available low-resolution images, the goodness of image registration and so on [5]. To resolve this, some novel methods were proposed. For example, the example-based methods assume that the lost high-frequency information is available in other natural images, which can be covered by the exemplar low resolution and high resolution patch pairs [6] [7]; the sparse representation based method utilizes the idea that “the sparse representation of a low-resolution image patch can be applied with the high-resolution image patch dictionary to generate a high-resolution image patch” [8]. However, both methods require large numbers of images for training. For sparse representation based method, how to select the patch size is still an open question.

The performances of the upscaling algorithms or super resolution algorithms have been studied subjectively for HD or lower resolutions [9] [10]. However, few studies have been conducted on the evaluation of the upscaling algorithms for UHD resolution. In this study, we aim at investigating the acceptability of the up-scaled UHD compared to the original UHD video sequences. In other words, we want to know the difference in terms of QoE between the up-scaled UHD versions and the native UHD version, to conclude

This work has been financially supported by the UltraHD-4U project.

on the necessity of research on novel upscaling algorithms for UHD.

The structure of this paper is shown as follows. Section 2 introduces the subjective experiment, including the video sequences used in the test, the apparatus, the test environment, the viewers, the test methodology and the test procedure. Section 3 provides the experimental results. Section 4 concludes the paper.

2. EXPERIMENT

2.1. Source video contents (SRC)

Eight video sequences were selected in this study. The thumbnails of these sequences are shown in Fig. 1. “Fountain”, “Runner” and “Tall Building” are from SJTU 4K Video Sequences [11]. “Table Car” and “Birthday” are provided by Technicolor. “Wiener” is provided by MediAventures. “Tears of Steel (ToS)” is from Mango Open Movie Project (mango.blender.org). We extracted two clips from “ToS” which are named as “ToS2” and “ToS3”.

These sequences were selected based on the idea that “in which conditions the observers’ viewing experience would best benefit from 4K resolution”. Thus, the sequences with high frequency component and high luminance contrast are chosen. For example, “Fountain” and “Tall Building” are with high frequency components and limited motion component, thus, observers may be attracted by the small droplets in “Fountain” and detailed building structure in “Tall Building”. “Table Car” and “Birthday” are with high color contrast and appear sharper than other sequences. “ToS2” and “Runners” contain large amount of local motion and high frequency objects, e.g., laser grid in “ToS2” and tree leaves in “Runners”. “Wiener” and “ToS3” are with pure black background and high textured object with local motion, i.e., a woman is playing with a gun in “Wiener” and the subtitle is scrolling in “ToS3”.

These sequences were originally with the resolution of 4K (3840×2160). To test the upscaling algorithms, we down-scaled these sequences to 1080p and 720p with the JSVM (Joint Scalable Video Model) software (which performed the best in a preliminary test with different down-scaling algorithms).

These sequences are originally with different frame rates. In this test, we play back the videos frame by frame with 25fps (thus, movements would appear a bit slower than in reality). We did not using temporal downsampling methods because they often introduce more artifacts than this slowdown effect, in particular if the ratio is not an integer value. The length of these sequences are 10 seconds.

Spatial perceptual information (SI) and temporal perceptual information (TI), as described in ITU-T recommendation P.910 [12], are calculated and shown in Fig. 2.

2.2. Upscaling algorithms

In this test, the upscaling algorithms are considered as the HRC(Hypothetical Reference Circuit). The upscaling algorithms were selected based on the criterions that 1) Source code is available; 2) Performance is acceptable; 3) Computational cost is feasible.

We have tested several upscaling algorithms, i.e., bilinear, bicubic, sinc, lanczos, Catmull-Rom [13], Mitchell-Netravali [14], Pocs (*Projections onto convex sets*) [15], RSR (*Robust Super Resolution*) [16] and ScSR (*Sparse-coding based Super Resolution*) [8], etc. Pretest results showed that as low complexity algorithms, bilinear performs the worst and sinc is too similar to lanczos, so bilinear and sinc are excluded. Lanczos-3 is the best compromise between accuracy and ringing artifacts, so it has been selected over Lanczos-2 and Lanczos-4. For Catmull-Rom and

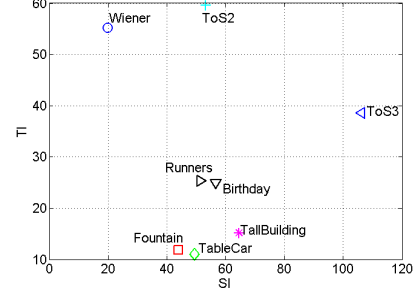


Fig. 2. Spatial Information (SI) and Temporal Information (TI) of the video sequences.

Mitchell-Netravali, the standard definition of these two filters is:

$$k(x) = \frac{1}{6} \begin{cases} (12 - 9B - 6C)|x|^3 + (-18 + 12B + 6C)|x|^2 + (6 - 2B) & \text{if } |x| < 1 \\ (-B - 6C)|x|^3 + (6B + 30C)|x|^2 + (-12B - 48C)|x| + (8B + 24C) & \text{if } 1 \leq |x| < 2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

According to the pretest results, we set $B=0$ and $C=0.5$ for Catmull-Rom filter, and $B=C=\frac{1}{3}$ for Mitchell-Netravali filter for optimum performance. The Pocs and RSR are selected due to their acceptable performance and feasible time complexity. ScSR is excluded due to its noisy results on objects edges. In addition, we added a HRC condition, which is a combination of lanczos-3 and an additional sharpening processing, to see the influence of sharpening on UHD.

As there are two resolutions (720p and 1080p) for each video content, but only a limited number can be tested in the subjective experiment, we decided to split the SRC into two groups according to the characteristics of the content as explained in section 2.1 and shown in Fig. 1. For Group 1, we applied the upscaling algorithms on 1080p versions of SRC. For Group 2, we applied the upscaling algorithms on 720p versions of SRC. Theoretically, the upscaled sequences from 720p should generally be less preferred by observers than the upscaled sequences from 1080p. To make a comparison between the general “1080p upscaling” condition and the best “720p upscaling” condition, and the opposite way, we added 720p lanczos-3 to Group 1 and 1080p RSR to Group 2. The original 4K resolution sequences are included in the HRCs as well to be a reference. The HRC details can be found in Table I.

2.3. Experimental Setup

The 4K display used in the test is Panasonic TX-L65WT600E with the size of 65 inches. In [17] the authors pointed out that realism is best perceived in 0.75H (H - Height of the screen) viewing distance. However, according to the discrimination power of a normal vision observer, which is on average one minute of arc corresponding to a critical pattern frequency of 30 cycles per degree (cpd), the viewing distance should be 1.6H. In this study, we decided to set the viewing distance to 1H (84cm) as a compromise between immersion effects and visual acuity.

The test environment is set as a standard subjective quality test condition according to ITU-R BT.500 [18], as shown in Fig.3.



Fig. 1. Source video contents in the test.

Table I. HRC list

HRC	Up-scaling algorithm	original size	SRC Group1	SRC Group2
1	Reference	-	√	√
2	Bicubic	720p		√
3		1080p	√	
4	Lanczos3	720p	√	√
5		1080p	√	
6	Catmull-Rom	720p		√
7		1080p	√	
8	Mitchell-Netravali	720p		√
9		1080p	√	
10	Lanczos3 +sharpen30	720p		√
11		1080p	√	
12	Pocs	720p		√
13		1080p	√	
14	RSR(Robust Super Resolution)	720p		√
15		1080p	√	√

2.4. Viewers

Forty-two viewers participated in this subjective experiment. Eighteen are male, twenty-four are female. They are all non-expert in subjective experiment, image processing or 4K related field. Their ages ranged from 19 to 68 with average age of 32. All have either normal or corrected-to-normal visual acuity. The visual acuity test was conducted with a Snellen Chart for near vision which is particularly necessary for the UHD quality assessment task as the viewing distance is only 84cm. The Ishihara plates were used for color vision test. All of the viewers passed the pre-experiment vision check.

2.5. Assessment Methodology: Pair Comparison

Pair Comparison test methodology was selected in the test due to its simplicity and accuracy. Moreover, it is considered as the “most reliable” solution for the assessment of multi-dimensional QoE [19] [20]. The basic idea of pair comparison is to show two video sequences at one time, and ask the observers to select which one they prefer. In our test, observers only compare the video sequences which have the same video content but different HRCs. To display one pair on one screen, we split the screen into 8 vertical stripes and each sequence of the pair occupies one stripe alternatively as shown



Fig. 3. Test environment

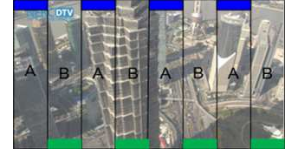


Fig. 4. Presentation method.

d_1	d_2	d_3
d_8	d_9	d_4
d_7	d_6	d_5

Fig. 5. Arrangement of matrix \mathbf{R}_{ARD} .

in Fig.4. To uniquely identify the two sequences, we put blue and green labels on the upper and lower position of the screen, and a small vertical shift was added to emphasize the correspondence to one particular HRC. The question for each observer is “which one do you prefer?”. Observers could select the green one or the blue one by the corresponding button on a touchpad.

To reduce the number of test pairs for each observer, the Adaptive Rectangular Design (ARD) [20] was used in the test. We implemented the ARD method in such a way:

1. For the first observer, the indices of the 9 HRCs were randomly placed into a square matrix \mathbf{R} of size 3×3 . The comparison is between pair $\{S_i S_j\}$ if and only if $(i, j) \in \text{set } \mathbf{C}$, where \mathbf{C} is defined as: $\mathbf{C} = \{(x, y) | p = p' \vee q = q'\}$ where $\{x = r_{pq}, y = r_{p'q'} \text{ in } \mathbf{R}\}$.
2. According to the obtained paired comparison data from all observers who have conducted the test, the Bradley-Terry scores (see Section 3.1) and the rank orders of all stimuli can be obtained. Supposing that the ascending ordering index vector for all stimuli is $\mathbf{d} = (d_1, d_2, \dots, d_9)$, the square matrix \mathbf{R}_{ARD} is arranged in such a way that the elements of the vector \mathbf{d} are placed along a spiral as shown in Fig. 5.
3. For the next observer $k(k > 1)$, based on the obtained \mathbf{R}_{ARD} , comparisons are conducted on the pairs which satisfy the conditions as described in step 1 (replacing \mathbf{R} by \mathbf{R}_{ARD}).
4. Repeat step 2 and 3, until all observers finished the test.

In such way, a 3×3 matrix would generate 18 pairs. 8 SRCs would generate in total 124 pairs for each observer.

To avoid memory effects from observers, we constrain the randomization of the presentation order of the video sequences as follows:

- The presentation of the sequence content is as random as possible, no observer watches the same content in two consecutive presentations.
- For each observer, the presentation order for each sequence is balanced, i.e., $\{S_i S_j\}$, $\{S_k S_i\}$. This means for all the pairs which include sequence S_i , half of the pairs should show S_i on the odd stripes, the rest should show S_i on the even stripes.
- For all observers, all the pairs of sequences must be displayed in both orders. For example, if one observer watches $\{S_i S_j\}$, there must be another observer who watches $\{S_j S_i\}$. To ensure this, the pairs for observer $2K - 1$ and observer $2K$ ($K > 0$) are exactly the same except for the $\{S_i S_j\}$ presentation order. \mathbf{d} and \mathbf{R}_{ARD} are updated not for each observer but each second observer, e.g., they are updated after observer $2K$ finishing the test, afterwards, the playlist for observer $2K + 1$ and $2K + 2$ is generated.

2.6. Procedure

The subjective experiment contained a training session and a test session. In the training session, there were five pairs of video sequences. At the beginning, the observers were told that they would watch a series of 4K videos. The stripes marked with same color are from the same source. The observers should select which source they prefer by pressing the corresponding “blue” or “green” button. During the training session, all questions of the viewers were answered. We ensured that after the training session, all of the viewers knew about the process and task of this experiment clearly.

In the main test session, the task and procedure were the same as the training session except 124 pairs of video sequences were compared. To avoid visual fatigue caused by long time watching affecting the experimental results, the viewers were asked to have a 5 minutes break after half of the test samples. The whole test was approximately 40 minutes.

3. RESULTS

3.1. Observers’ Preference of Experience (PoE)

The outcome of the paired comparison test is a pair comparison matrix \mathbf{A} , where $\mathbf{A} = (a_{ij})_{m \times m}$. a_{ij} is the total count of preference of stimulus S_i over S_j for all observers. $a_{ii} = 0$ for $i = 1, 2, \dots, m$. The total number of comparisons for video pair $\{S_i S_j\}$ is $n_{ij} = a_{ij} + a_{ji}$.

The Bradley-Terry (BT) model [21] [22] is a well-known model to convert pair comparison data to psychophysical scale values for all video sequences. It is defined as follows:

$$V_i - V_j = \log \frac{P_{ij}}{1 - P_{ij}} \quad (2)$$

P_{ij} represents the probability that video S_i is preferred to S_j . P_{ij} can be estimated by p_{ij} , where $p_{ij} = a_{ij}/n_{ij}$ when n_{ij} is large enough. The outputs are the differences of the BT scale values between stimuli S_i and S_j , i.e., $V_i - V_j$. By utilizing the least squares estimation or the maximum likelihood estimation, the scale value V_i for each stimulus, $i = 1, \dots, m$ can be estimated. Please

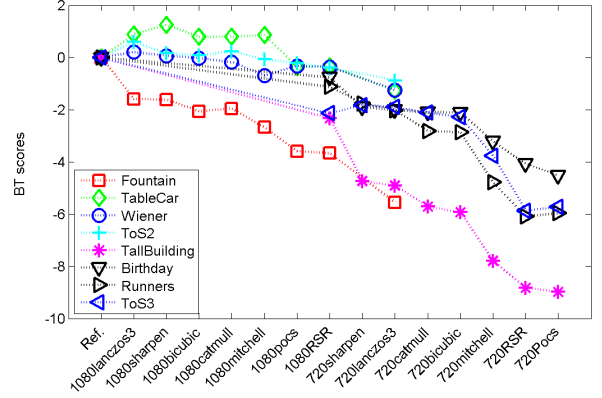


Fig. 6. BT scores for all sequences.

note that the scale value V_i is a relative value which can be added with an offset, but cannot be re-scaled by a factor. In this study, BT scores represent observer’s preference. The higher the BT scores, the higher PoE. Particularly, for convenient comparison, we set the BT scores of the reference sequences to 0.

The BT scores of all video sequences are shown in Fig. 6. For better visualization of the confidence intervals and rank order of the HRCs for each video content, the individual video content BT scores are shown in Fig. 7. It should be noted that the confidence intervals of BT score could not be used as a reference of significance test as it is influenced by both goodness of model fit and the total number of comparisons.

As shown in Fig. 6, generally, the original 4K sequences perform the best, then follows the up-scaled version from 1080p, the up-scaled version from 720p is the worst. Considering the upscaling algorithms, lanczos-3 and lanczos-3 with sharpening perform the best, then follows bicubic, Catmull and Mitchell. The much more complex super resolution algorithm Pocs and RSR perform the worst.

The performance of the upscaling algorithms is also content dependent. As shown in Fig.6, the reference is not the best in “Table Car”, “Wiener” and “ToS2”. According to the confidence intervals, we may infer that observers’ certainty on the preference on different upscaled “Wiener” and “ToS2” may not be as significant as in other sequences.

3.2. Significance test on PoE

To test the significance of the observers’ preference on each pair, the Barnard’s exact test [23] [20] was used here. As the pairs were updated for the observers, there might be some pairs which were never compared or the number of comparisons is too small for statistic test, in such case, only the pairs whose comparison number is larger than 10 are considered in the Barnard’s test. The test results are shown in Table II.

As shown in the table, in most cases, observers’ preference on different up-scaled video sequences is significant. In particular, the number of significantly different preference on 720p upscaled sequences is much higher than on 1080p, which means observers perceptual sensitivity on 720p upscaled sequences is higher. In addition, the results showed that generally the classic and

Table II. Barnard test on each pair. * represents significant difference. - represents there is not enough number of comparisons (<10). > represents the observers prefer row element to column element. < represents the observers prefer column element to row element. 1 represents 1080p, 7 represents 720p.

	Ref	1bic	1lanc	1cat	1mit	1shp	1Pocs	1RSR	7bic	7lanc	7cat	7mit	7shp	7Pocs	7RSR
Ref	-	<	>	>	>*	>*	>*	>*	>*	>*	>*	>*	>*	>*	>*
1bic	>	-	<	<	>	<	>*	>*	-	>*	-	-	-	-	-
1lanc	<	>	-	>	>*	<	>	>*	-	>*	-	-	-	-	-
1cat	<	>	<	-	>	<	>*	>	-	>*	-	-	-	-	-
1mit	<*	<	<*	<	-	<*	>*	>	-	>*	-	-	-	-	-
1shp	<*	>	>	>	>*	-	>*	>*	-	>*	-	-	-	-	-
1Pocs	<*	<*	<	<*	<*	<*	-	>	-	>*	-	-	-	-	-
1RSR	<*	<*	<*	<	<	<*	<	-	>*	>*	>*	>*	>*	>*	>*
7bic	<*	-	-	-	-	-	-	<*	-	<*	<	>*	<*	>*	>*
7lanc	<*	<*	<*	<*	<*	<*	<*	<*	>*	-	>*	>*	<	>*	>*
7cat	<*	-	-	-	-	-	-	<*	>	<*	-	>*	<*	>*	>*
7mit	<*	-	-	-	-	-	-	<*	<*	<*	<*	-	<*	>*	>*
7shp	<*	-	-	-	-	-	-	<*	>*	>	>*	>*	-	>*	>*
7Pocs	<*	-	-	-	-	-	-	<*	<*	<*	<*	<*	<*	-	<
7RSR	<*	-	-	-	-	-	-	<*	<*	<*	<*	<*	<*	>	-

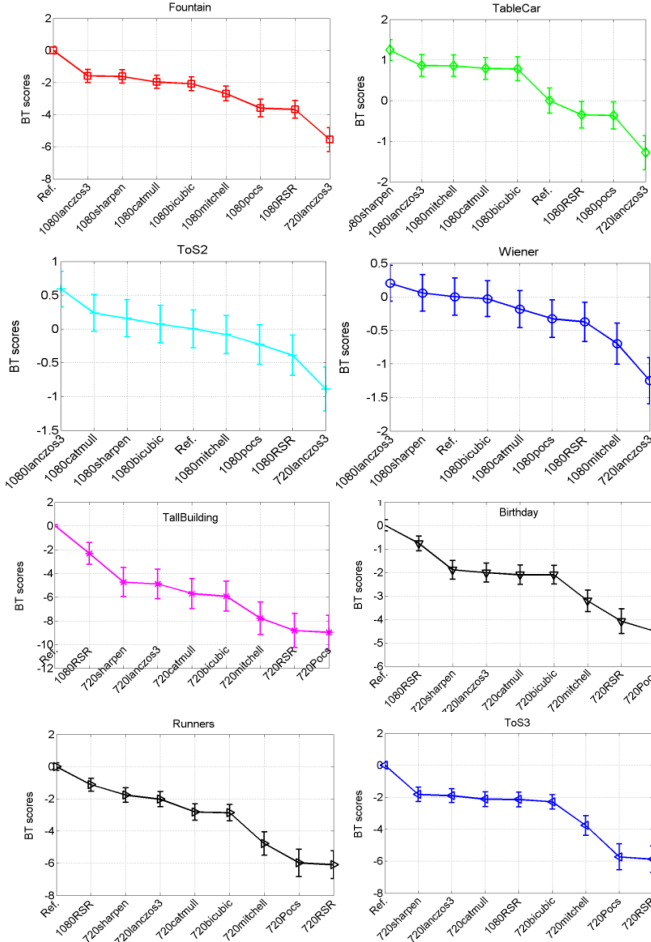


Fig. 7. The sorted BT scores for each video content.

computationally inexpensive algorithms, i.e., bicubic, lanczos-3, Catmull, Mitchell, and sharpening are preferred to Pocs and RSR, which is consistent with the results of BT scores.

3.3. Discussion

The test results indicated that the performance of the upscaling algorithms to 4K resolution is content dependent: 1) In “Wiener” and “ToS2”, as the video contains large motion activity, observers may find it difficult to distinguish between different upscaling algorithms. As shown in Fig. 7, the confidence intervals of different upscaling algorithms are large and overlapping. This is expected because the human visual acuity decreases when object motion is fast; 2) For the case of sequences containing high frequency components and slow motion activity, e.g., “ToS3”, “Tall Building”, “Birthday” and “Fountain”, observers gained benefit from 4K resolution as the reference provides better experience; 3) For the case of sequences containing large flat areas with moderate brightness, e.g., “Table car”, the camera noise may be perceived easily and the upscaling algorithm would reduce or blur this artifact. Thus, upscaled content is experienced better than the reference.

It is more important to note that there are some particularly different mechanisms of human visual system work on watching UHD TV. For example, when watching UHD TV in a shorter distance, the visual field has been increased. In our study, the visual angle for viewers is 53 degree vertically and 83 degree horizontally. The highest resolution for the human vision is reached in the visual field of 15 degree for vertical and 20 degree for horizontal, then decreases from 15-30 degree, and finally peripheral vision processes are achieved at angles larger than 30 degree. Thus, in our case, peripheral vision would play a greater role which may lead to the following observations: 1) Observer’s visual acuity on high motion content on 4K screen is significantly lower than in the normal HD condition; 2) As the peripheral vision is much more sensitive to flickering than the fovea, jerkiness effects would be more perceptible; 3) Center bias might be more significant due to

the comfortable visual field since watching the peripheral content would need extra effort. In addition, high resolution and large screen induced “immersive” feeling may also generate motion sickness due to the illusion of self-motion. All the above mentioned criteria should be considered as the influence factors of viewing experience of UHD.

According to the test results in this study, in general conditions, the high complexity super resolution algorithms perform worse than the classic and lower computational complexity algorithms. As we already mentioned before, the mechanisms of human visual system on watching UHDTV is significantly different from HDTV, thus, perceptual preference on different types of distortion would be different, an example is that sharpening is quite preferred in the test. Therefore, an UHDTV-oriented upscaling algorithm considering visual field, center bias and peripheral vision may improve QoE of UHDTV.

4. CONCLUSIONS

In this study, we have evaluated the performance of several upscaling algorithms by upscaling the video sequences from 720p, 1080p to 4K resolution. The results indicated that in general conditions, the current state of the art upscaling algorithms could not achieve as good perceptual quality as the original UHD versions. In addition, the best upscaling algorithm may be not the state-of-the-art computationally expensive super resolution algorithms but the less costly one, for example, lanczos-3 eventually with added sharpening. Due to the different viewing conditions on UHD and the corresponding viewing behavior of observers, improvement may be expected if the upscaling algorithm is particularly adapted to UHD.

5. ACKNOWLEDGMENT

The participants of the subjective experiment are gratefully acknowledged.

6. REFERENCES

- [1] P. Le Callet, S. Möller, and A. Perkis, “Qualinet white paper on definitions of quality of experience v.1.1,” *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Jun. 2012.
- [2] R. Keys, “Cubic convolution interpolation for digital image processing,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [3] C. E. Duchon, “Lanczos filtering in one and two dimensions,” *Journal of Applied Meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979.
- [4] K. Turkowski, “Filters for common resampling tasks,” in *Graphics gems*. Academic Press Professional, Inc., 1990, pp. 147–165.
- [5] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, “Joint map registration and high-resolution image estimation using a sequence of undersampled images,” *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [6] W. T. Freeman, T.R. Jones, and E. C. Pasztor, “Example-based super-resolution,” *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.
- [7] T. A. Stephenson and T. Chen, “Adaptive markov random fields for example-based super-resolution of faces,” *EURASIP Journal on Advances in Signal Processing*, 2006.
- [8] J.C. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [9] A.R. Reibman and T. Schaper, “Subjective performance evaluation of super-resolution image enhancement,” *Second Int. Wkshp on Video Proc. and Qual. Metrics (VPQM)*, 2006.
- [10] Y. Pitrey, M. Barkowsky, and R. and others Le Callet, P. and Pepion, “Subjective Quality Evaluation of H. 264 High-Definition Video Coding versus Spatial Up-Scaling and Interlacing,” *QoE for Multimedia Content Sharing*, 2010.
- [11] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, “The SJTU 4K video sequence dataset,” in *the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013)*, Klagenfurt, Austria, 2013.
- [12] ITU-T P.910, “Subjective video quality assessment methods for multimedia applications,” *International Telecommunication Union*, Apr. 2008.
- [13] E. Catmull and R. Rom, “A class of local interpolating splines,” *Computer aided geometric design*, vol. 74, pp. 317–326, 1974.
- [14] D. P. Mitchell and A. N. Netravali, “Reconstruction filters in computer-graphics,” in *ACM Siggraph Computer Graphics*. ACM, 1988, vol. 22, pp. 221–228.
- [15] M.K. Ozkan, A. M. Tekalp, and M. I. Sezan, “Pocs-based restoration of space-varying blurred images,” *Image Processing, IEEE Transactions on*, vol. 3, no. 4, pp. 450–454, 1994.
- [16] A. Zomet, A. Rav-Acha, and S. Peleg, “Robust super-resolution,” in *Proceedings of the international conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2001, vol. 1, pp. I–645.
- [17] K. Masaoka, M. Emoto, M. Sugawara, and F. Okano, “Presence and preferable viewing conditions when using an ultra high-definition large-screen display,” in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 28–37.
- [18] ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” *International Telecommunication Union*, Geneva, Switzerland, Jan. 2012.
- [19] J-S. Lee, L. Goldmann, and T. Ebrahimi, “Paired comparison-based subjective quality assessment of stereoscopic images,” *Multimedia Tools and Applications*, pp. 1–18, Feb. 2012.
- [20] J. Li, M. Barkowsky, and P. Le Callet, “Subjective assessment methodology for preference of experience in 3DTV,” in *IVMSP Workshop, 2013 IEEE 11th*. IEEE, 2013, pp. 1–4.
- [21] R.A. Bradley, “14 paired comparisons: Some basic procedures and examples,” *Handbook of Statistics*, vol. 4, pp. 299–326, 1984.
- [22] R.A. Bradley and M.E. Terry, “Rank analysis of incomplete block designs: I. The method of paired comparisons,” *Biometrika*, vol. 39, no. 3/4, pp. 324–345, Dec. 1952.
- [23] G. Barnard, “A new test for 2×2 tables,” *Nature*, vol. 156:177, 1945.